Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 18, No. 3, pp. 529–543, 2022

A DESCENT FAMILY OF THREE–TERM CONJUGATE GRADIENT METHODS WITH GLOBAL CONVERGENCE FOR GENERAL FUNCTIONS*

Maryam Khoshsimaye-Bargard and Ali Ashrafi†

Abstract: In this paper, we introduce a family of three–term conjugate gradient methods for solving unconstrained optimization problems in which the conjugate parameter satisfies a restrictive relation $|\beta_k| \leq \beta_k^{FR}$. Also, the sufficient descent property holds when the line search fulfills the strong Wolfe line search. The global convergence of the presented method is proved under the strong Wolfe line search with some mild conditions and even without convexity assumption on the objective function. Numerical experiments are performed on a set of test problems of the CUTEr library, the results of which illustrate the practical effectiveness of this method.

Key words: *unconstrained optimization, three–term conjugate gradient method, sufficient descent condition, global convergence*

Mathematics Subject Classification: *49M37, 90C06, 90C30*

1 Introduction

An unconstrained optimization problem can generally be formulated as

$$\min_{x\in\mathbb{R}^n} f(x), \tag{1.1}$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ is smooth and its gradient at point x is denoted by $g(x)$.

One of the most efficient iterative methods for solving problem (1.1) is the conjugate gradient (CG) method, which produces a sequence of iterations such as $\{x_k\}_{k\geq 0}$ with the rule

$$x_0 \in \mathbb{R}^n,\ x_{k+1} = x_k + \alpha_k d_k,\ k \geq 0, \tag{1.2}$$

in which d_k is the search direction defined by

$$d_0 = -g_0,\ d_{k+1} = -g_{k+1} + \beta_k d_k,\ k \geq 0, \tag{1.3}$$

where $g_k = g(x_k) = \nabla f(x_k)$, β_k in (1.3) is called the conjugate parameter and the positive scalar α_k in (1.2) is known as the step length, which is typically determined by some line

*This research was supported by the Research Council of Semnan University.
†Corresponding author.

search techniques. Different selections for β_k lead to generate different CG methods. Among them there exist the Dai–Yuan method [11], the Fletcher–Reeves (FR) method [16], the Hestenes–Stiefel method [21], the Polak–Ribière–Polyak (PRP) method [32, 33] as well as their various generalizations see for example [2, 3, 6, 15, 20].

The step length α_k is the solution of the one–dimensional minimization problem known as the line search. Exact line search is the most common type of line search which is expressed as

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(x_k + \alpha d_k).$$

This technique is computationally expensive because of the need to calculate the accurate value of the step length. So, we usually utilize the inexact line search. The strong Wolfe line search is one of the most famous inexact line searches, in which α_k satisfies

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k \bigtriangledown f(x_k)^T d_k, \tag{1.4}$$

$$| \bigtriangledown f(x_k + \alpha_k d_k)^T d_k| \leq -\sigma \bigtriangledown f(x_k)^T d_k, \tag{1.5}$$

while in the Wolfe line search requires α_k to satisfy (1.4) and

$$\bigtriangledown f(x_k + \alpha_k d_k)^T d_k \geq \sigma \bigtriangledown f(x_k)^T d_k, \tag{1.6}$$

where $0 < \delta < \sigma < 1$.

The method introduced by Fletcher and Reeves [16] with the conjugate parameter

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \tag{1.7}$$

in which $\|.\|$ denotes the Euclidean norm, is considered as one of the most well–known classical CG methods. This method has good characteristics such as the termination property of the algorithm for finite quadratic functions and the global convergence feature. Hence, the global convergence analysis of the FR method has been extensively studied. As a cursory glance, Zoutendijk [39] first created the convergence results for the FR method with the exact line search. Afterwards, Al–Baali [1] proved the global convergence of this method for general objective functions using the strong Wolfe conditions (1.4) and (1.5) with $\sigma < \frac{1}{2}$. Later, Liu et al. [28] developed the convergence results presented by Al–Baali for $\sigma = \frac{1}{2}$. Gilbert and Nocedal presented a comprehensive study on the global convergence of the FR method in [17].

In spite of the theoretical merits and the strong convergence features of the FR method, the numerical performance of this method is fundamentally influenced by jamming phenomenon [17, 34]. So, the improvement of the FR method from computationally aspect has attracted special attentions. For instance, based on the spectral gradient method presented by Barzilai and Borwein [8], Zhang et al. [37] suggested a modification of the classical FR method called the spectral FR conjugate gradient method with the search direction

$$d_0 = -g_0, \ d_{k+1} = -\theta_k g_{k+1} + \beta_k d_k, \ k \geq 0, \tag{1.8}$$

where $\beta_k = \beta_k^{FR}$ and θ_k is known as the spectral parameter and calculated by

$$\theta_k = \frac{d_k^T y_k}{\|g_k\|^2}, \tag{1.9}$$

in which $y_k = g_{k+1} - g_k$. An important and appealing property of this method is that at each iteration, the produced search direction independent of any line search fulfills in the sufficient descent condition, namely

$$g_k^T d_k \leq -\tau \|g_k\|^2, \ \forall k \geq 0, \tag{1.10}$$

with constant $\tau > 0$. It is obvious that if the line search is exact, then the method becomes the classical FR conjugate gradient method. Recently, Liu et al. [26] introduced a family of spectral CG methods with the spectral parameter

$$\theta_k = -\frac{g_k^T d_k}{\|g_k\|^2} + \beta_k \frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2}, \tag{1.11}$$

where its search direction always satisfies in the sufficient descent condition (1.10). They also proved that their proposed method is globally convergent with the Wolfe line search when $|\beta_k| \leq \beta_k^{FR}$. It is worth noting that some other well-known conjugate parameters with the property $|\beta_k| \leq \beta_k^{FR}$ have been introduced, see for example [17, 22, 35]. Lately, Li and Cao [24] suggested another conjugate parameter with limitation $|\beta_k| \leq \beta_k^{FR}$ as

$$\beta_k^{DPRP} = \frac{g_{k+1}^T y_k}{\eta_k \|g_k\|^2}, \tag{1.12}$$

in which

$$\eta_k = 1 + \mu \frac{|g_{k+1}^T g_k|}{\|g_{k+1}\|^2},$$

and $\mu \geq 1$ is a constant.

Guaranteeing the sufficient descent condition for the classical CG methods is difficult because of the low number of parameters. Hence, the modification of them to produce methods with more parameters and efficiency has been extensively studied. Among these modifications, the three–term CG method, first introduced by Beale [9] in 1972, has attracted the attention of many researchers. Three–term CG methods are another important class of CG methods designed to increment the efficiency of classical CG methods. Numerical results illustrate that three–term CG algorithms are more efficient, robust and reliable compared to classical CG algorithms [4, 38]. Interested readers can refer to more resources on three–term CG methods [5, 7, 14, 23, 25, 27, 29].

The theoretical and numerical merits of the three–term CG methods encourage us to offer another family of three–term CG methods for solving large–scale optimization problems. In this regard, motivated by argument provided in [26], we try to calculate the value of the coefficient of the third term in such a way that the sufficient descent condition holds for its search direction independent of the convexity of the objective function. Moreover, in order to ensure the global convergence of the proposed method using the strong Wolfe line search, we consider the conjugate parameter such that the relation $|\beta_k| \leq \beta_k^{FR}$ holds.

The remainder of this manuscript is arranged as follows. In Section 2, we suggest a specific class of three–term CG methods that possesses sufficient descent property. The global convergence analysis of our proposed method under some mild conditions is presented in Section 3, and the results of our numerical experiments on a set of unconstrained optimization test problems are made in Section 4. Eventually, in Section 5 we provide the conclusion.

2 A Family of Three–Term Conjugate Gradient Methods

In this paper, based on the approach of [26], we suggest a family of three–term CG methods to solve problem (1.1), which the structure of its search direction is as follows:

$$d_0 = -g_0, \; d_{k+1} = -g_{k+1} + \beta_k d_k - \nu_k y_k, \quad k \geq 0. \tag{2.1}$$

Multiplying both sides of the above equation by g_{k+1}^T, we obtain

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -\|g_{k+1}\|^2 + \beta_k g_{k+1}^T d_k - \nu_k g_{k+1}^T y_k \\ &= \|g_{k+1}\|^2 \frac{g_k^T d_k}{\|g_k\|^2} \left(-\frac{\|g_k\|^2}{g_k^T d_k} + \beta_k \frac{\|g_k\|^2}{g_k^T d_k} \frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2} - \nu_k \frac{\|g_k\|^2}{g_k^T d_k} \frac{g_{k+1}^T y_k}{\|g_{k+1}\|^2} \right), \end{aligned}$$

which by definition

$$\delta_{k+1} = -\frac{\|g_k\|^2}{g_k^T d_k} + \beta_k \frac{\|g_k\|^2}{g_k^T d_k} \frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2} - \nu_k \frac{\|g_k\|^2}{g_k^T d_k} \frac{g_{k+1}^T y_k}{\|g_{k+1}\|^2},$$

it follows that

$$\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} = \frac{g_k^T d_k}{\|g_k\|^2} \delta_{k+1}. \tag{2.2}$$

It is clear that if $\delta_{k+1} \equiv 1$ holds for each $k \geq 0$, then relations (2.1) and (2.2) imply

$$\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} = \frac{g_k^T d_k}{\|g_k\|^2} = \cdots = \frac{g_0^T d_0}{\|g_0\|^2} = -1,$$

Hence, we have

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2, \quad \forall k \geq 0.$$

This means that if we can determine ν_k such that δ_{k+1} satisfies $\delta_{k+1} \equiv 1$, then the sufficient descent condition always holds for the search direction d_{k+1} independent of any line search.

According to the above analysis and after some algebraic manipulations, we can express ν_k as

$$\nu_k = \beta_k \frac{g_{k+1}^T d_k}{g_{k+1}^T y_k}, \quad k \geq 0.$$

To ensure the convergence of our suggested method, we modify the parameter ν_k as follows:

$$\nu_k = \begin{cases} \beta_k \frac{g_{k+1}^T d_k}{g_{k+1}^T y_k}, & g_{k+1}^T g_k \leq 0, \\ 0, & otherwise. \end{cases} \tag{2.3}$$

Now, we present the algorithm of the proposed method as follows:

Algorithm 2.1. (TTCG method)

Step 0. Choose the initial point $x_0 \in \mathbb{R}^n$, and the scalars $\varepsilon > 0$ and $0 < \delta < \sigma < \frac{1}{2}$. Calculate $f_0 = f(x_0)$ and $g_0 = \nabla f(x_0)$. Set $d_0 = -g_0$ and $k := 0$.

Step 1. If $\|g_k\|_\infty < \varepsilon$, then stop.

Step 2. Compute the step length α_k by the strong Wolfe line search conditions (1.4) and (1.5).

Step 3. Define β_k and calculate ν_k satisfying (2.3). Then compute the next search direction d_{k+1} by (2.1).

Step 4. Generate the next iterate by $x_{k+1} = x_k + \alpha_k d_k$.

Step 5. Set $k := k + 1$, and **goto** Step 1.

3 Global Convergence Property

In this section, we illustrate that if the conjugate parameter β_k is suitably bounded in magnitude, then the Algorithm 2.1 is globally convergent. So throughout this section, we assume that β_k is any scalar such that

$$| \beta_k | \leq \beta_k^{FR}, \quad \forall k \geq 0. \tag{3.1}$$

The following standard assumptions are also required in our analysis.

Assumption 3.1. The level set $\mathcal{L}_0 = \{x \in \mathbb{R}^n | f(x) \leq f(x_0)\}$ is bounded.

Assumption 3.2. In the neighborhood $\mathcal{N}$ of $\mathcal{L}_0$, the objective function f is continuously differentiable, and its gradient is Lipschitz continuous; that is, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathcal{N}. \tag{3.2}$$

Under the above Assumptions, there exist positive constants B and γ such that

$$\|x - y\| \leq B, \quad \forall x, y \in \mathcal{L}_0,$$

and

$$\|g(x)\| \leq \gamma, \quad \forall x \in \mathcal{L}_0. \tag{3.3}$$

The following lemma shows that the proposed method satisfies the condition (1.10).

Lemma 3.3. *Suppose that Assumptions 3.1 and 3.2 hold. The search direction defined by* (2.1)*, in which ν_k is computed by* (2.3) *under the strong Wolfe line search conditions* (1.4) *and* (1.5) *with $\sigma \in (0, \frac{1}{2})$, fulfills the sufficient descent condition* (1.10) *with $\tau = \frac{1-2\sigma}{1-\sigma}$.*

Proof. Based on the search direction defined by (2.1) with ν_k calculated by (2.3), we must consider the following two cases:

Case (*i*): $g_{k+1}^T g_k \leq 0$. From the discussion presented in the previous section, we know that

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2, \quad \forall k \geq 0,$$

hence, we have

$$g_{k+1}^T d_{k+1} \leq -\tau\|g_{k+1}\|^2, \quad \forall k \geq 0.$$

Case (*ii*): $g_{k+1}^T g_k > 0$. In this situation, following carefully the proof of Lemma 3.1 of [17], we first prove by induction that the search direction satisfies

$$-\frac{1}{1-\sigma} \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq \frac{2\sigma - 1}{1-\sigma}. \tag{3.4}$$

For this purpose, it is clear that the above relation holds for $k = 0$, since the value of the intermediate expression is equal to -1. Now we suppose that (3.4) is true for $k \geq 0$, which implies $g_k^T d_k < 0$, because

$$\frac{2\sigma - 1}{1 - \sigma} < 0.$$

On the other hand, we have

$$\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} = -1 + \beta_k \frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2} = -1 + \frac{\beta_k}{\beta_k^{FR}} \frac{g_{k+1}^T d_k}{\|g_k\|^2}. \tag{3.5}$$

It also follows from the strong Wolfe line search condition (1.5) that

$$|\beta_k g_{k+1}^T d_k| \leq -\sigma |\beta_k| g_k^T d_k.$$

Combining this with (3.5), we obtain

$$-1 + \sigma \frac{|\beta_k|}{\beta_k^{FR}} \frac{g_k^T d_k}{\|g_k\|^2} \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -1 - \sigma \frac{|\beta_k|}{\beta_k^{FR}} \frac{g_k^T d_k}{\|g_k\|^2},$$

which using the left side relation of the induction hypothesis we have

$$-1 - \frac{|\beta_k|}{\beta_k^{FR}} \frac{\sigma}{1 - \sigma} \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -1 + \frac{|\beta_k|}{\beta_k^{FR}} \frac{\sigma}{1 - \sigma}.$$

From $|\beta_k| \leq \beta_k^{FR}$, we conclude that (3.4) holds for $k+1$. Now with the definition

$$\tau = \frac{1 - 2\sigma}{1 - \sigma},$$

we get

$$g_{k+1}^T d_{k+1} \leq -\tau \|g_{k+1}\|^2.$$

□

The following lemma has an useful conclusion called the Zoutendijk condition, which is frequently used to prove the global convergence of CG methods.

Lemma 3.4. [12] *Suppose that x_0 is a starting point for which Assumptions 3.1 and 3.2 hold. Consider the itreative method* (1.2)*, where d_k is a descent direction and α_k satisfies* (1.4) *and* (1.6)*. Then we have that*

$$\sum_{k=0}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} < \infty. \tag{3.6}$$

Lemma 3.5. [10] *Suppose that Assumptions 3.1 and 3.2 hold. Consider any conjugate gradient method in the form* (1.2) *and* (1.3)*, where d_k is a descent direction and α_k is obtained by the strong Wolfe line search. If*

$$\sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} = \infty, \tag{3.7}$$

we have that

$$\liminf_{k \to \infty} \|g_k\| = 0. \tag{3.8}$$

Now, we demonstrate that our proposed method for general functions is globally convergent.

Theorem 3.6. *If Assumptions 3.1 and 3.2 hold, then Algorithm 2.1, under strong Wolfe conditons* (1.4) *and* (1.5) *with* $\sigma \in (0, \frac{1}{2})$*, converges in the sense that* (3.8) *holds.*

Proof. Assume by contradiction that conclusion (3.8) is incorrect. Then there is a constant $\varepsilon > 0$, such that

$$\|g_k\| \geq \varepsilon, \quad \forall k \geq 0. \tag{3.9}$$

According to the structure of the search direction calculated by (2.1) with ν_k determined by (2.3), we have to consider the following two situations:

Case (*i*): $g_{k+1}^T g_k \leq 0$. In this case, from (2.1) we have

$$\begin{aligned} d_{k+1} &= -g_{k+1} + \beta_k d_k - \beta_k \frac{g_{k+1}^T d_k}{g_{k+1}^T y_k} y_k \\ &= -g_{k+1} + \beta_k \left(I - \frac{y_k g_{k+1}^T}{g_{k+1}^T y_k} \right) d_k. \end{aligned}$$

Taking into account Lemma 1.1 of [30], the relations (3.1), (3.2), (3.3) and (3.9),we get

$$\begin{aligned} \|d_{k+1}\| &\leq \|g_{k+1}\| + |\beta_k| \left\| I - \frac{y_k g_{k+1}^T}{g_{k+1}^T y_k} \right\| \|d_k\| \\ &\leq \|g_{k+1}\| + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \frac{\|y_k\| \|g_{k+1}\|}{|g_{k+1}^T y_k|} \|d_k\| \\ &= \|g_{k+1}\| + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \frac{\|y_k\| \|g_{k+1}\|}{\left| \|g_{k+1}\|^2 - g_{k+1}^T g_k \right|} \|d_k\| \\ &\leq \|g_{k+1}\| + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \frac{\|y_k\| \|g_{k+1}\|}{\|g_{k+1}\|^2} \|d_k\| \\ &\leq \gamma + \frac{\gamma L \alpha_k \|d_k\|}{\varepsilon^2} \|d_k\|, \end{aligned}$$

where similar to the proof of Lemma 3.1 of [36], since $\alpha_k d_k \to 0$ as $k \to \infty$, there exist a constant $\zeta \in (0,1)$ and an integer k_0, such that the following inequality holds for all $k \geq k_0$:

$$\frac{\gamma L}{\varepsilon^2} \alpha_k \|d_k\| \leq \zeta.$$

Hence, for each $k > k_0$, we have

$$\begin{aligned} \|d_{k+1}\| \leq \gamma + \zeta \|d_k\| &\leq \gamma + \zeta \big(\gamma + \zeta \|d_{k-1}\| \big) \\ &\leq \gamma \left(1 + \zeta + \zeta^2 + \cdots + \zeta^{k-k_0} \right) + \zeta^{k-k_0+1} \|d_{k_0}\| \\ &\leq \frac{\gamma}{1-\zeta} + \|d_{k_0}\|. \end{aligned}$$

By setting

$$C = \max \left\{ \|d_0\|, \|d_1\|, \cdots, \|d_{k_0}\|, \frac{\gamma}{1-\zeta} + \|d_{k_0}\| \right\},$$

we get $\|d_k\| \leq C$ for all $k \geq 0$, and consequently

$$\sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} \geq \sum_{k=0}^{\infty} \frac{1}{C^2} = \infty,$$

where from Lemma 3.5 follows

$$\liminf_{k\to\infty} \|g_k\| = 0,$$

while it contradicts (3.9). Hence, the proof is complete.

Case (*ii*): $g_{k+1}^T g_k > 0$. In this situation, by rewriting equation (1.3) as

$$\beta_k d_k = d_{k+1} + g_{k+1},$$

and squaring its sides, we get

$$\beta_k^2 \ \|d_k\|^2 = \|d_{k+1}\|^2 + \|g_{k+1}\|^2 + 2d_{k+1}^T g_{k+1},$$

where using Lemma 3.3, we have

$$\begin{aligned}\beta_k^2 \ \|d_k\|^2 &\leq \|d_{k+1}\|^2 + \|g_{k+1}\|^2 - 2\tau\|g_{k+1}\|^2 \\ &\leq \|d_{k+1}\|^2 + \|g_{k+1}\|^2.\end{aligned}$$

Therefore, we can write

$$\|d_{k+1}\|^2 \geq \ \beta_k^2 \ \|d_k\|^2 - \|g_{k+1}\|^2. \tag{3.10}$$

On the other hand, by multiplying the equation (1.3) by g_{k+1}^T, we obtain

$$\begin{aligned}\|g_{k+1}\|^2 &= \beta_k \ g_{k+1}^T d_k - g_{k+1}^T d_{k+1} \\ &\leq |\beta_k||g_{k+1}^T d_k| + |g_{k+1}^T d_{k+1}|.\end{aligned} \tag{3.11}$$

Since the step length α_k satisfies the strong Wolfe conditions, it follows that

$$|g_{k+1}^T d_k| \leq \sigma|g_k^T d_k|. \tag{3.12}$$

Hence, the inequalities (3.11) and (3.12) imply that

$$\|g_{k+1}\|^2 \leq \sigma|\beta_k||g_k^T d_k| + |g_{k+1}^T d_{k+1}|. \tag{3.13}$$

Using general inequality

$$(a_1b_1 + a_2b_2 + \cdots + a_nb_n)^2 \leq \left(a_1^2 + a_2^2 + \cdots + a_n^2\right)\left(b_1^2 + b_2^2 + \cdots + b_n^2\right),$$

where

$$a_1 = \sigma \ , \ a_2 = 1 \ , \ b_1 = |\beta_k||g_k^T d_k| \ , \ b_2 = |g_{k+1}^T d_{k+1}| \ ,$$

the inequality (3.13) becomes

$$\begin{aligned}\|g_{k+1}\|^4 &\leq \left(\sigma|\beta_k||g_k^T d_k| + |g_{k+1}^T d_{k+1}|\right)^2 \\ &\leq \left(\sigma^2 + 1\right)\left(\beta_k^2(g_k^T d_k)^2 + (g_{k+1}^T d_{k+1})^2\right).\end{aligned}$$

Now, according to positiveness of $c = \frac{1}{1+\sigma^2}$, we can write

$$c\|g_{k+1}\|^4 \leq \beta_k^2(g_k^T d_k)^2 + (g_{k+1}^T d_{k+1})^2. \tag{3.14}$$

Dividing the sides of the inequality (3.10) by $\|d_k\|^2$, we obtain

$$\frac{\|d_{k+1}\|^2}{\|d_k\|^2} - \beta_k^2 \geq -\frac{\|g_{k+1}\|^2}{\|d_k\|^2}. \tag{3.15}$$

On the other hand, we know

$$\frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} + \frac{(g_k^T d_k)^2}{\|d_k\|^2} = \frac{1}{\|d_{k+1}\|^2}\left((g_{k+1}^T d_{k+1})^2 + \frac{\|d_{k+1}\|^2}{\|d_k\|^2}(g_k^T d_k)^2\right). \tag{3.16}$$

Now, adding and subtracting the sentence $\beta_k^2\left(g_k^T d_k\right)^2$ to the right-hand side of the relation (3.16) leads to

$$\begin{aligned}
&\frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} + \frac{(g_k^T d_k)^2}{\|d_k\|^2} \\
&= \frac{1}{\|d_{k+1}\|^2}\left((g_{k+1}^T d_{k+1})^2 + \beta_k^2(g_k^T d_k)^2 + \frac{\|d_{k+1}\|^2}{\|d_k\|^2}(g_k^T d_k)^2 - \beta_k^2(g_k^T d_k)^2\right) \qquad (3.17)\\
&= \frac{1}{\|d_{k+1}\|^2}\left((g_{k+1}^T d_{k+1})^2 + \beta_k^2(g_k^T d_k)^2 + \left(\frac{\|d_{k+1}\|^2}{\|d_k\|^2} - \beta_k^2\right)(g_k^T d_k)^2\right). \qquad (3.18)
\end{aligned}$$

By substituting (3.14) and (3.15) in the equation (3.17), we get

$$\frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} + \frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \frac{1}{\|d_{k+1}\|^2}\left(c\|g_{k+1}\|^4 - \frac{\|g_{k+1}\|^2}{\|d_k\|^2}(g_k^T d_k)^2\right).$$

Therefore, we can write

$$\frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} + \frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2}\left(c - \frac{(g_k^T d_k)^2}{\|d_k\|^2\|g_{k+1}\|^2}\right).$$

Taking into account the Zoutendijk condition (3.6) and the relation (3.9), we have $\lim_{k\to\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} = 0$, which yields for large enough k

$$\frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} + \frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2}\left(\frac{1}{1+\sigma^2}\right).$$

So, by Lemma 3.4 and convergence criterion of series, we get

$$\sum_{k=0}^{\infty} \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} < \infty,$$

and consequently

$$\sum_{k=0}^{\infty} \frac{1}{\|d_{k+1}\|^2} \leq \frac{1}{\varepsilon^4}\sum_{k=0}^{\infty} \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} < \infty,$$

which is inconsistent with (3.7) and completes the proof. □

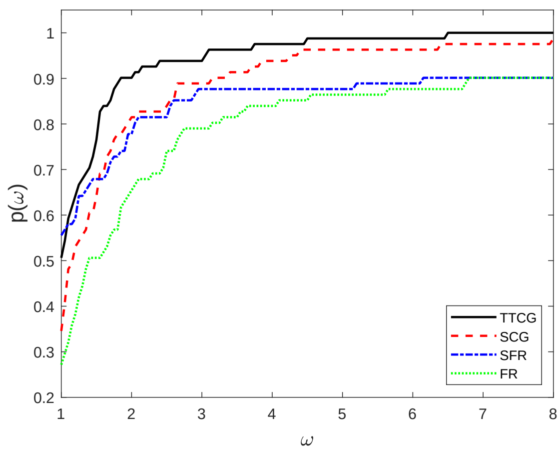

Figure 1: Total number of function and gradient evaluations performance profiles

4 Numerical Experiments

In this section, the computational efficiency of the TTCG method with β_k defined by (1.12) is evaluated. Here we set $\mu = 3.5$ in (1.12), due to the booming numerical results obtained between the various values of $\mu \in \{0.5k\}_{k=2}^{20}$.

The TTCG method is compared with the following three CG methods, with the search direction calculated by (1.8) in which β_k and θ_k are respectively computed by:

FR: (1.7) and $\theta_k = 1$;

SFR: (1.7) and (1.9);

SCG: (1.12) and (1.11).

All algorithms have been tested on 81 functions of the CUTEr collection [18] with the minimum dimension being equal to 50, as given in Table 1, applying MATLAB 7.7.0.471 (R2008b) installed on a computer Intel(R) Core(TM) i5-8250U CPU @ 1.80 GHz with 8.00 GB RAM and the Centos 6.2 server Linux operation system. The algorithms were terminated by achieving a maximum of 10000 iterations or reaching a solution with $\|g_k\|_\infty < 10^{-6}\,(1 + |f(x_k)|)$.

In addition, efficiency comparisons were drawn using the Dolan–Moré performance profile [13], on the running time (CPUT) and the total number of function and gradient evaluations (TNFGE) being equal to $N_f + 3N_g$, where N_f and N_g respectively signify the number of function and gradient evaluations [19]. The performance profile gives, for every $\omega \geq 1$, the proportion $p(\omega)$ of the test problems whereas each considered algorithmic variant has performance within a factor of ω of the best.

All calculated algorithms are performed by the Wolfe line search conditions (1.4) and (1.6) with $\delta = 0.0001$ and $\sigma = 0.3$ base on Algorithm 3.5 of [31].

Figures 1 and 2 exhibit the comparison results of the above methods. As can be seen, the TTCG method is superior to other methods in terms of the total number of function and gradient evaluations, and the CPU time. So, the above numerical experiments illustrate

Table 1: Outputs

No. function	n	TTCG		SCG		SFR		FR	
		TNFGE	CPUT	TNFGE	CPUT	TNFGE	CPUT	TNFGE	CPUT
1. ARGLINA	200	12	0.1490	12	0.0543	12	0.0554	12	0.0545
2. BDEXP	5000	16	0.0641	16	0.0475	16	0.0461	16	0.0544
3. BIGGSB1	5000	32526	3.1800	60769	6.0069	40018	4.8102	40363	4.7317
4. BQPGABIM	50	306	0.0249	313	0.0261	374	0.0269	511	0.0313
5. BQPGASIM	50	306	0.0220	313	0.0215	374	0.0269	511	0.0318
6. BROWNAL	200	89	0.0498	89	0.0442	147	0.0474	793	0.0983
7. BRYBND	5000	224	0.1280	237	0.1272	429	0.1763	390	0.1656
8. CHNROSNB	50	6525	0.2310	7146	0.2487	40030	1.3783	44116	1.4246
9. CLPLATEB	5041	60270	14.6000	60117	14.8550	40098	10.2030	43119	10.7720
10. COSINE	10000	167	0.1770	693	0.4886	315	0.2578	109	0.1318
11. CRAGGLVY	5000	431	0.3010	1339	0.8209	526	0.3588	697	0.4652
12. CURLY10	10000	907	0.3470	921	0.3465	40089	11.9720	43707	13.1770
13. CURLY20	10000	1888	0.9470	3604	1.6981	40087	18.8840	19535	8.9403
14. CURLY30	10000	4635	2.7100	6286	3.6225	40098	25.2030	43097	25.4590
15. DECONVU	63	49934	1.7300	61132	2.0852	7686	0.2863	51686	1.7143
16. DIXMAANA	3000	36	0.0453	36	0.0364	51	0.0391	56	0.0404
17. DIXMAANB	3000	32	0.0417	32	0.0382	37	0.0383	37	0.0392
18. DIXMAANC	3000	37	0.0440	37	0.0367	37	0.0398	37	0.0424
19. DIXMAAND	3000	42	0.0434	42	0.0374	42	0.0361	42	0.0358
20. DIXMAANE	3000	3032	0.3860	2897	0.3641	1697	0.2383	3854	0.4776
21. DIXMAANF	3000	2164	0.2930	3979	0.4913	1624	0.2304	2958	0.3779
22. DIXMAANG	3000	1660	0.2270	2402	0.3093	1441	0.2076	3773	0.4752
23. DIXMAANH	3000	1964	0.2710	2029	0.2725	1248	0.1878	4550	0.5615
24. DIXMAANI	3000	15034	1.8400	24276	5.1125	19573	2.4071	40485	4.8116
25. DIXMAANJ	3000	1368	0.1990	3566	0.4417	2249	0.3061	4499	0.5544
26. DIXMAANK	3000	3116	0.3930	2099	0.2751	1697	0.2377	3109	0.3930
27. DIXMAANL	3000	1676	0.2270	1650	0.2210	1516	0.2171	2741	0.3501
28. DIXON3DQ	10000	41085	6.4600	60950	9.8790	40010	7.0644	40030	6.7440
29. DMN15103	99	81661	187.0000	80987	188.9600	40576	104.1400	54827	133.2100
30. DMN37142	66	60068	118.0000	60334	120.3200	46970	96.5750	46702	95.4930
31. DMN37143	99	74605	172.0000	75672	172.5200	42865	107.4000	54929	133.3700
32. DQDRTIC	5000	617	0.1640	765	0.1847	446	0.1308	1083	0.2334
33. DQRTIC	5000	4	0.0325	4	0.0321	4	0.0257	4	0.0332
34. DRCAV1LQ	4489	4	0.0608	4	0.0656	4	0.0660	4	0.0662
35. DRCAV2LQ	4489	4	0.0605	4	0.0651	4	0.0638	4	0.0652
36. DRCAV3LQ	4489	4	0.0600	4	0.0636	4	0.0648	4	0.0650
37. EDENSCH	2000	127	0.0509	137	0.0464	218	0.0566	170	0.0497
38. EG2	1000	23	0.0204	23	0.0216	23	0.0301	23	0.0193
39. EIGENALS	2550	60176	86.6000	60678	89.7420	40049	62.1010	48259	70.9810
40. EIGENBLS	2550	59146	87.4000	59342	89.1190	40049	64.8150	43470	70.2470
41. EIGENCLS	2652	59754	94.8000	59875	94.3630	40050	65.3340	44949	71.3020
42. ENGVAL1	5000	80	0.0711	80	0.0686	145	0.0803	199	0.0909
43. EXTROSNB	1000	60894	3.3400	60380	3.1626	16381	0.8985	51447	2.5904
44. FLETCBV2	5000	4	0.0636	4	0.0568	4	0.0611	4	0.0618
45. FLETCHBV	5000	148	0.1260	148	0.1190	148	0.1176	148	0.1177
46. FLETCHCR	1000	5175	0.3410	6889	0.4420	4266	0.2968	5927	0.3797
47. FMINSRF2	5625	3924	0.8510	4660	1.0055	40022	8.9954	41942	9.1008
48. FMINSURF	5625	4828	1.0900	5599	1.2565	40022	9.1349	42201	9.3909
49. FREUROTH	5000	342	0.1570	316	0.1370	241	0.1221	454	0.1714
50. GENHUMPS	5000	4	0.0447	4	0.0398	4	0.0410	4	0.0410
51. GENHUMPS	5000	4	0.0321	4	0.0369	4	0.0446	4	0.0327
52. GENROSE	500	15815	0.7650	16014	0.7645	40033	1.9336	43471	1.9934
53. LIARWHD	5000	1839	0.3450	2285	0.4122	1106	0.2256	1318	0.2594
54. MANCINO	100	100	0.1690	100	0.1669	100	0.1636	95	0.1590
55. MOREBV	5000	700	0.1480	629	0.1365	1631	0.2889	2211	0.3670
56. MSQRTALS	1024	39083	11.7000	60322	18.4660	22787	7.4139	41123	13.2610
57. MSQRTBLS	1024	37562	11.9000	60239	19.1650	15808	5.3120	41199	13.4510
58. NCB20	5010	485	0.3940	487	0.3993	40211	24.7180	45665	27.2410
59. NCB20B	5000	401	0.3470	443	0.3584	398	0.3313	483	0.3801
60. NONCVXU2	5000	4	0.0294	4	0.0364	4	0.0288	4	0.0306
61. NONDIA	5000	2218	0.3590	4054	0.6225	495	0.1095	637	0.1297
62. PENALTY1	1000	1151	0.0759	792	0.0576	4091	0.2355	3586	0.1959
63. PENALTY2	200	4	0.0028	4	0.0150	4	0.0042	4	0.0109
64. POWER	10000	7101	0.9300	9705	1.3748	2296	0.3644	2632	0.3934
65. QUARTC	5000	4	0.0315	4	0.0305	4	0.0295	4	0.0255
66. SCHMVETT	5000	97	0.1350	80	0.1179	162	0.1888	198	0.2202
67. SENSORS	100	150	0.2990	106	0.2210	4934	8.1966	1745	2.9258
68. SINQUAD	5000	776	0.4830	973	0.5450	41463	16.7800	369	0.2861
69. SPARSINE	5000	60433	24.4000	60576	24.0140	40028	16.2610	40727	16.4190
70. SPARSQUR	10000	166	0.1540	231	0.1715	318	0.2117	327	0.2114
71. SPMSRTLS	4999	2220	0.5690	2555	0.6490	3048	0.7678	4458	1.0506
72. SROSENBR	5000	274	0.0660	925	0.1284	557	0.0903	367	0.0709
73. TESTQUAD	5000	61039	6.4400	60824	6.7074	40087	4.0486	44598	4.1488
74. TOINTGOR	50	713	0.0410	804	0.0416	595	0.0356	1146	0.0541
75. TOINTGSS	5000	71	0.0756	66	0.0701	127	0.0994	370	0.2177
76. TOINTPSP	50	951	0.0415	887	0.0431	1842	0.0699	1190	0.0497
77. TOINTQOR	50	172	0.0195	168	0.0198	281	0.0253	416	0.0242
78. TRIDIA	5000	22149	2.1500	51919	4.9870	20382	2.1518	42256	4.2726
79. VARDIM	200	8	0.0020	8	0.0103	8	0.0015	8	0.0083
80. VAREIGVL	50	140	0.0206	148	0.0206	412	0.0314	573	0.0314
81. WOODS	4000	1983	0.2580	2877	0.3548	1676	0.2231	658	0.1094

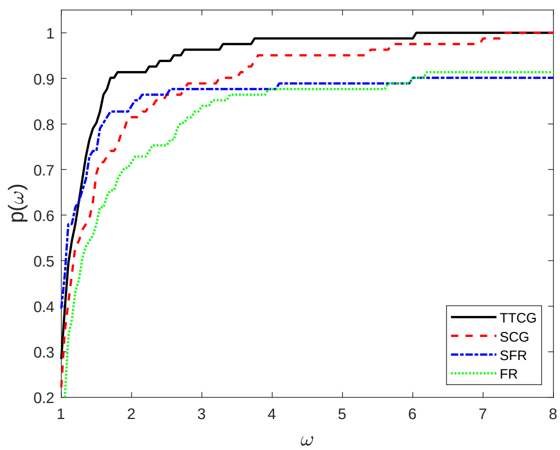

Figure 2: CPU time performance profiles

that the suggested method is promising. Our experiments showed that averagely in 81.54% of the iterations of TTCG method we had $\nu_k = \beta_k \frac{g_{k+1}^T d_k}{g_{k+1}^T y_k}$.

5 Conclusion

In this paper, following the successful approach presented by Liu et al. [26], we introduced a family of three–term conjugate gradient methods in such a way that the sufficient descent condition holds. Convergence analysis for general functions has been provided by the strong Wolfe line search under proper conditions when $|\beta_k| \leq \beta_k^{FR}$. Moreover, our suggested method has been numerically compared with some existing effective methods on a set of 81 unconstrained optimization test problems of the CUTEr library. The results illustrated that our method is promising and efficient based on the Dolan–Moré performance profile.

Acknowledgment

The authors thank the anonymous reviewers for their valuable comments and suggestions helped to improve the quality of this work.

References

[1] M. Al–Baali, Descent property and global convergence of the Fletcher–Reeves method with inexact line search, *IMA J. Numer. Anal.* 5 (1985) 121–124.

[2] Z. Aminifard and S. Babaie–Kafaki, A modified descent Polak–Ribière–Polyak conjugate gradient method with global convergence property for nonconvex functions, *Calcolo* 56 (2019) 1–11.

[3] Z. Aminifard and S. Babaie–Kafaki, Matrix analyses on the Dai–Liao conjugate gradient method, *ANZIAM J.* 61 (2019) 195–203.

[4] N. Andrei, A new three–term conjugate gradient algorithm for unconstrained optimization, *Numer. Algorithms* 68 (2015) 305–321.

[5] N. Andrei, A simple three–term conjugate gradient algorithm for unconstrained optimization, *J. Comput. Appl. Math.* 241 (2013) 19–29.

[6] S. Babaie–Kafaki, A quadratic hybridization of Polak–Ribière–Polyak and Fletcher–Reeves conjugate gradient methods, *J. Optim. Theory Appl.* 154 (2012) 916–932.

[7] S. Babaie–Kafaki and R. Ghanbari, Two modified three–term conjugate gradient methods with sufficient descent property, *Optim. Lett.* 8 (2014) 2285–2297.

[8] J. Barzilai and J.M. Borwein, Two–point step size gradient methods. *IMA J. Numer. Anal.* 8 (1988) 141–148.

[9] E.M.L. Beale, A derivation of conjugate gradients, in: *Numerical Methods for nonlinear Optimization,* Academic Press, New York, 1972.

[10] Y.H. Dai and L.Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods, *Appl. Math. Optim.* 43 (2001) 87–101.

[11] Y.H. Dai and Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, *SIAM J. Optim.* 10 (1999) 177–182.

[12] Y.H. Dai and Y.X. Yuan, Convergence of three–term conjugate gradient methods, *Chinese J. Numer. Math.* 21 (1999) 69–78.

[13] E.D. Dolan and J.J. Mor 辿, Benchmarking optimization software with performance profiles, *Math. Program. Ser. A,* 91 (2002) 201–213.

[14] X. Dong, D. Han, Z. Dai, L. Li, and J. Zhu, An accelerated three–term conjugate gradient method with sufficient descent condition and conjugacy condition, *J. Optim. Theory Appl.* 179 (2018) 944–961.

[15] X. Dong, H. Liu, and Y. He, A self–adjusting conjugate gradient method with sufficient descent condition and conjugacy condition, *J. Optim. Theory Appl.* 165 (2015) 225–241.

[16] R. Fletcher and C.M. Reeves, Function minimization by conjugate gradients, *Comput. J.* 7 (1964) 149–154.

[17] J.C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, *SIAM J. optim.* 2 (1992) 21–42.

[18] N.I. Gould, D. Orban, and P.L. Toint, CUTEr and SifDec: A constrained and unconstrained testing environment, revisited, *ACM Trans. Math. Software* 29 (2003) 373–394.

[19] W.W. Hager and H. Zhang, Algorithm 851: CG–DESCENT, a conjugate gradient method with guaranteed descent, *ACM Trans. Math. Software* 32 (2006) 113–137.

[20] W.W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods, *Pacific J. Optim.* 2 (2006) 335–358.

[21] M.R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. Nat. Bur. Stand.* 49 (1952) 409–436.

[22] Y.F. Hu and C. Storey, Global convergence result for conjugate gradient methods, *J. Optim. Theory Appl.* 71 (1991) 399–405.

[23] M. Li, A family of three–term nonlinear conjugate gradient methods close to the memoryless BFGS method, *Optim. Lett.* 12 (2018) 1911–1927.

[24] C. Li and X. Cao, A Three–Team Conjugate Gradient Method with Disturbance Factors for Unconstrained Optimization, *Journal of Physics: Conference Series.* 1168 (2019) 1–5.

[25] X. Li, J. Shi, X. Dong, and J. Yu, A new conjugate gradient method based on quasi–newton equation for unconstrained optimization, *J. Comput. Appl. Math.* 350 (2019) 372–379.

[26] J.K. Liu, Y.M. Feng, and L.M. Zou, A spectral conjugate gradient method for solving large–scale unconstrained optimization, *Comput. Math. Appl.* 77 (2019) 731–739.

[27] J.K. Liu, Y.M. Feng, and L.M. Zou, Some three–term conjugate gradient methods with the inexact line search condition, *Calcolo* 55 (2018) 1–16.

[28] G.H. Liu, J.Y. Han, and H.X. Yin, Global convergence of the Fletcher–Reeves algorithm with inexact linesearch, *Appl. Math. J. Chinese Univ. Ser. B.* 10 (1995) 75–82.

[29] J.K. Liu, Y.X. Zhao, and X.L. Wu, Some three–term conjugate gradient methods with the new direction structure, *Appl. Numer. Math.* 150 (2020) 433–443.

[30] S. Nataj and S.H. Lui, Superlinear convergence of nonlinear conjugate gradient method and scaled memoryless BFGS method based on assumptions about the initial point, *Appl. Math. Comput.* 369 (2020) 1–15.

[31] J. Nocedal and S.J. Wright, *Numerical Optimization,* Springer, New York, 2006.

[32] E. Polak and G. Ribière, Note sur la convergence de méthodes de directions conjuguées, *Rev. Fr. Imform. Rech. Oper.* 3 (1969) 35–43.

[33] B.T. Polyak, The conjugate gradient method in extremal problems, *USSR Comp. Math. Math. Phys.* 9 (1969) 94–112.

[34] M.J.D. Powell, Restart procedures for the conjugate gradient method, *Math. program.* 12 (1977) 241–254.

[35] D. Touati–Ahmed and C. Storey, Efficient hybrid conjugate gradient techniques, *J. Optim. Theory Appl.* 64 (1990) 379–397.

[36] L. Zhang, W. Zhou, and D. Li, A descent modified Polak–Ribière–Polyak conjugate gradient method and its global convergence, *IMA J. Numer. Anal.* 26 (2006) 629–640.

[37] L. Zhang, W. Zhou, and D. Li, Global convergence of a modified Fletcher–Reeves conjugate gradient method with Armijo–type line search, *Numer. Math.* 104 (2006) 561–572.

[38] L. Zhang, W. Zhou, and D. Li, Some descent three–term conjugate gradient methods and their global convergence, *Optim. Methods Softw.* 22 (2007) 697–711.

[39] G. Zoutendijk, Nonlinear programming, computational methods, in: *Integer and Nonlinear Programming*, J. Abadie (ed)., North–Holland, Amsterdam, 1970.

Manuscript received 22 June 2021
revised 20 August 2021
accepted for publication 16 September 2021

Maryam Khoshsimaye-Bargard

Department of Mathematics, Faculty of Mathematics
Statistics and Computer Science
Semnan University, Semnan, Iran
E-mail address: m.khoshsima@semnan.ac.ir

Ali Ashrafi
Department of Mathematics, Faculty of Mathematics
Statistics and Computer Science
Semnan University, P.O. Box: 35131-19111, Semnan, Iran
E-mail address: a_ashrafi@semnan.ac.ir